University of Zurich

Socioeconomic Institute
Sozialökonomisches Institut

Working Paper No. 1012

A Causal Interpretation of Extensive and Intensive Margin Effects in Generalized Tobit Models

Kevin E. Staub

November 2010

Socioeconomic Institute
University of Zurich

Working Paper No. 1012

A Causal Interpretation of Extensive and Intensive Margin Effects in Generalized Tobit Models

November 2010

Author's address: Kevin E. Staub
E.mail: kevin.staub@econ.uzh.ch

Publisher Sozialökonomisches Institut
Bibliothek (Working Paper)
Rämistrasse 71
CH-8006 Zürich
Phone: +41-44-634 21 37
Fax: +41-44-634 49 82
URL: www.soi.uzh.ch
E-mail: soilib@soi.uzh.ch

A causal interpretation of extensive and intensive margin effects in generalized Tobit models*

Kevin E. Staub**

Department of Economics, University of Zurich

November 2, 2010

Abstract: The usual decomposition of effects in corner solution models into extensive and intensive margins is generally incompatible with a causal interpretation. This paper proposes a decomposition based on the joint distribution of potential outcomes which is meaningful in a causal sense. The difference between decompositions can be substantial and yield diametrically opposed results, as shown in a standard Tobit model example. In a generalized Tobit application exploring the effect of reducing firm entry regulation on bilateral trade flows between countries, estimates suggest that using the usual decomposition would overstate the contribution of the extensive margin by around 15%.

Keywords: Limited dependent variables, potential outcomes, causality, conditional-on-positives effect, Tobit, two-part model, country margins of trade.

JEL Codes: C24, C34, F14.

*An earlier version of this paper circulated under the title "A causal interpretation of extensive and intensive margin effects in corner solution models".

Acknowledgements: I am indebted to Rainer Winkelmann for extensive discussions and many helpful suggestions which improved this article. I would like to thank Colin Cameron for an insightful discussion, as well as Gregori Baetschmann and Stefan Boes for useful inputs. I am also grateful to conference participants at the European Trade Study Group Conference in Fribourg, Switzerland, and the Midwest Econometrics Group Annual Meeting in St. Louis, MO, USA, as well as seminar participants at the University of Zurich, for various comments. Opinions and errors are mine.

** ✉ University of Zurich, Department of Economics, Chair for Statistics and Empirical Economic Research, Zürichbergstr. 14, CH-8032 Zürich, Switzerland, ☎ +41 44 634 2312, ✉ kevin.staub@econ.uzh.ch

1 Introduction

Many outcomes of interest in economics are nonnegative and have a cluster of observations at the value zero. Prominent examples include working hours, health care demand, and expenditure data. More generally, variables with these features are referred to as corner solution outcomes (Wooldridge, 2002), which suggests the idea of utility maximization under constraints where both interior and corner solutions occur, for instance due to kinks in budget constraints.

Researchers analyzing effects of variables on corner solution outcomes frequently take interest in decomposing the effect into the part attributable to individuals starting to participate (called *extensive margin*), and the part attributable to already participating individuals (called *intensive margin*). The decomposition used is algebraically straightforward as it is based on factoring the expectation of the corner solution variable, say $\mathrm{E}(Y)$, into the participation probability $\Pr(Y > 0)$ and the conditional expectation $\mathrm{E}(Y|Y > 0)$ (McDonald and Moffitt, 1980). The extensive margin is driven by the participation effect [PE], the change in the probability to participate; the intensive margin is driven by the conditional-on-positives effect [COP], the change in the outcome given participation.

In contrast to the simplicity of the mechanical aspect, endowing the decomposition with a causal interpretation is substantially more problematic. For instance, recent work framing the problem in terms of Rubin's potential outcomes model has pointed out that COP effects do not measure the impact of a treatment on participating individuals; rather, they are hopelessly contaminated by a sort of selection bias, even in experimental settings (Angrist, 2001; Angrist and Pischke, 2009). An apparent solution is resorting to the interpretation of effects on underlying, latent variables such as in censored regression and sample-selection models, where causal interpretation is feasible. However, as these authors and others emphasize (cf. Dow and Norton, 2003), latent outcomes are artificial and lack a meaningful interpretation in corner solution contexts.

In this article, I propose a conceptually different decomposition of the effect into extensive and intensive margins. It is based on the joint distribution of potential outcomes, which ensures that the resulting parts are meaningful in a causal sense. Indeed, the new decomposition succeeds in representing the total effect as an average of the treatment effects for interesting subgroups of the population: those induced to participate by the treatment, and those participating regardless of it. Like the conventional decomposition, this one is not identified nonparametrically, although sharp bounds can be derived for the average treatment effect of the population subgroups. Imposing some more structure point-identifies the decomposition. Examples include a class of generalized Tobit models, which are widely used in applied research. For these models, the differences between decompositions can be major.

An application to the gravity model of trade compares the two decompositions in a real-world setting. The decomposition of trade effects into extensive and intensive margins is an issue of ongoing interest in the recent empirical trade literature (Felbermayr and Kohler, 2006; Helpman, Melitz and Rubinstein, 2008; Liu, 2009). Here, I estimate the effect of a hypothetical reduction in entry regulation costs on bilateral trade flows, and decompose it into country margins. The estimates suggest that the usual decomposition overstates the contribution of the extensive margin by around 15%.

Practitioners confronted with limited dependent variables in many diverse fields of applied economics and other social sciences will find this article to be of interest. Examples of work featuring the decomposition of effects into extensive and intensive margins include estimates of the effect of benefit-receiving on food expenditure (Hastings and Washington, 2010), the effect of various variables on intra- and inter-firm trade (Co, 2010), the effect of employer contributions on employee pension savings (Engelhardt and Kumar, 2007) the effect of worker productivity or unionization on working overtime (Sousa-Poza and Ziegler, 2003, and Trejo, 1993, respectively), the effect of managers' tax evasion preferences on underreporting corporate income (Joulfaian, 2000), the effect of various regressors on youth unemployment (Caspi et al. 1998), and the effect of health knowledge on health outcomes

(Kenkel, 1991). The list is neither complete nor representative, but it is suggestive of the widespread use of the decomposition in corner solution applications.

This article contributes to the growing recent literature on treatment effects for limited dependent variables (Aakvik, Heckman and Vytlacil, 2005; Chen, 2010; Chiburis, 2010; Fan and Wu, 2010). Since its emphasis lies on conceptual definition of objects of interest and interpretation, it is close in spirit to Angrist (2001). The representation of treatment effects as weighted sums of population groups is influenced by Angrist and Imbens (1994). As this framework is expressible in latent index models (Vytlacil, 2002), latent index representations with binary endogenous variables obtain expressions that resemble the ones presented below. Conceptually, they are quite different, because in the endogenous treatment literature, the population subgroups are defined by their potential treatment status in response to an instrument, while here the groups are defined by their potential outcome status in response to the (exogenous) treatment.

The plan for the article is this: the next section reprises the Angrist-Pischke "bad COP"-critique, presents the alternative decomposition and discusses its nonparametric identification. Section 3 exemplifies the new approach for some common Tobit-type models, and section 4 provides the application to the gravity model of trade. Section 5 contains a concluding discussion.

2 Corner solutions and potential outcomes

Consider the causal effect of a binary treatment T_i on the corner solution variable $Y_i \geq 0$ for individuals $i = 1, \ldots, N$. Let Y_{1i} denote the outcome for i if i received the treatment, i.e. $T_i = 1$, and Y_{0i} if $T_i = 0$, so that as usual

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})T_i$$

The focus here will be on the causal treatment effect $Y_{1i} - Y_{0i}$. Assume that the data comes from an ideal randomized controlled trial, so that assignment to treatment is random and compliance is perfect. Then, T_i is independent of (Y_{1i}, Y_{0i}), and the average treatment effect [ATE] $\mathrm{E}(Y_{1i} - Y_{0i})$ can be obtained from the prima-facie contrast $\mathrm{E}(Y_i|T = 1) - E(Y_i|T = 0)$. Using

$$\mathrm{E}(Y_i|T_i) = \Pr(Y_i > 0|T_i)\mathrm{E}(Y_i|Y_i > 0, T_i)$$

this contrast can be written as

$$\begin{aligned}
&\mathrm{E}(Y_i|T = 1) - E(Y_i|T_i = 0) = \\
&\quad \big\{ \Pr(Y_i > 0|T_i = 1) - \Pr(Y_i > 0|T_i = 0)\big\}\mathrm{E}(Y_i|Y_i > 0, T_i = 1) \\
&\quad +\big\{\mathrm{E}(Y_i|Y_i > 0, T_i = 1) - \mathrm{E}(Y_i|Y_i > 0, T_i = 0)\big\} \Pr(Y_i > 0|T_i = 0) \qquad (1)
\end{aligned}$$

This is the usual decomposition applied to limited dependent variables like Y_i in Tobit (Tobin, 1958) or Cragg (1971) models (McDonald and Moffitt, 1980; cf. also the standard graduate textbooks by Cameron and Trivedi, 2005, Greene, 2008, and Wooldridge, 2002). The first term after the equality sign is the extensive margin effect, which weights the PE —the term in curly brackets— by the expected Y_i conditional on participation; the second term is the intensive margin effect, which weights the COP (in curly brackets) by the probability of participation given $T_i = 0$. Angrist (2001) and Angrist and Pischke (2009) suggest rewriting COP in terms of potential outcomes as

$$\begin{aligned}
&\mathrm{E}(Y_i|Y_i > 0, T_i = 1) - \mathrm{E}(Y_i|Y_i > 0, T_i = 0) \\
&\quad = \mathrm{E}(Y_{1i}|Y_{1i} > 0, T_i = 1) - \mathrm{E}(Y_{0i}|Y_{0i} > 0, T_i = 0) \\
&\quad = \mathrm{E}(Y_{1i}|Y_{1i} > 0) - \mathrm{E}(Y_{0i}|Y_{0i} > 0) \\
&\quad = \mathrm{E}(Y_{1i} - Y_{0i}|Y_{1i} > 0) + \big\{\mathrm{E}(Y_{0i}|Y_{1i} > 0) - \mathrm{E}(Y_{0i}|Y_{0i} > 0)\big\} \qquad (2)
\end{aligned}$$

The second equality follows by independence of T_i from Y_{1i} and Y_{0i}. If only independence from Y_{0i} was assumed, as in Angrist's (2001) formulation, (2) would need to be written conditional on $T_i = 1$. This has no bearing on the present argument. As can be seen from the terms after the third equality, COP is composed of two terms. The first, $\mathrm{E}(Y_{1i} - Y_{0i}|Y_{1i} >$

0), can be given a causal interpretation. It is the treatment effect for the subpopulation of individuals having positive Y_i when $T_i = 1$. The second term, $\mathrm{E}(Y_{0i}|Y_{1i} > 0) - \mathrm{E}(Y_{0i}|Y_{0i} > 0)$, can be understood as a form of selection bias. The selection bias in COP arises because treatment has an effect on the composition of the group with $Y_i > 0$: The interest lies in those with $Y_{1i} > 0$, but the COP contrast also involves the group $Y_{0i} > 0$ which might be a super- or sub-set of the group $Y_{1i} > 0$, but not the same unless treatment has no effect on the participation probability. Thus, the analysis in (2) implies that using a decomposition like (1) cannot identify a causal effect even in ideal settings like a randomized controlled trial.

However, the more fundamental problem is that the first term in (2) is not an object of direct interest in a decomposition into extensive and intensive margins. The ATE for individuals with $Y_{1i} > 0$ mixes the ATE for the two population groups the decomposition set out to discriminate, the ones participating even without treatment and the ones participating because of the treatment.

2.1 *Decomposition based on joint outcomes*

Consider the following classification of individuals into non-overlapping and exhausting groups based on their *joint* distribution of potential outcomes, (Y_{0i}, Y_{1i}):

Group	Name	Potential outcomes
NP	Non-participants	$(Y_{0i} = 0, Y_{1i} = 0)$
S_1	Switchers	$(Y_{0i} = 0, Y_{1i} > 0)$
S_2	Switchers	$(Y_{0i} > 0, Y_{1i} = 0)$
P	Participants	$(Y_{0i} > 0, Y_{1i} > 0)$

Basing the definition of intensive and extensive margin effects on these groups clarifies their meaning substantially. The intensive margin effect is the contribution to the ATE of group P. Similarly, the extensive margin is the ATE contribution of switchers, i.e. those changing their participation status (groups S_1 and S_2). These are the objects of interest when decomposing causal effects into extensive and intensive margins; when researchers write about them, it is this what they mean (although they rarely state it so explicitly). For instance, take the labor economics example of working hours. The effect of a policy intervention increasing average working hours in the economy can be decomposed into

– the average change in hours worked of those working regardless of the intervention,

– plus the average hours worked by those joining the workforce because of the intervention,

– minus the average hours worked by those leaving the workforce because of the intervention,

the groups being weighted by their population fraction.

Often researchers choose models which possess some monotonicity assumption on the way treatment affects outcomes (Manski, 1997). This can lead to the elimination of one group out of S_1 and S_2. For instance, a strong monotone treatment response assumption states that the causal effect $Y_{1i} - Y_{0i}$ is either nonnegative or nonpositive for all i. In the working hours example, this means that if the policy increased working hours of workers, no one is induced to leave the workforce (group S_2 is ruled out). Such an assumption is embedded in the Tobit model. The monotone treatment response assumption can be weaker and still eliminate one group. Tautologically, it is sufficient that the causal effect is either positive for all i with $Y_{0i} = 0$ or $Y_{1i} = 0$, or negative. This assumption is implicit in Cragg's (1971) model. Often such assumptions are motivated by economic theory, and for many applications it might be plausible to impose them. Finally, the effect for individuals in group NP is always zero.

Thus, formally, the decomposition of the ATE based on the joint distribution of potential outcomes is

$$\begin{aligned} \mathrm{E}(Y_i|T=1) - \mathrm{E}(Y_i|T=0) &= \mathrm{E}_{Y_{1i},Y_{0i}}\left[\mathrm{E}(Y_{1i}-Y_{0i})|Y_{1i},Y_{0i}\right] \quad (3) \\ &= \mathrm{E}(Y_{1i}|Y_{0i}=0,Y_{1i}>0)\,\mathrm{Pr}(Y_{0i}=0,Y_{1i}>0) \\ &+ \mathrm{E}(-Y_{0i}|Y_{0i}>0,Y_{1i}=0)\,\mathrm{Pr}(Y_{0i}>0,Y_{1i}=0) \\ &+ \mathrm{E}(Y_{1i}-Y_{0i}|Y_{0i}>0,Y_{1i}>0)\,\mathrm{Pr}(Y_{0i}>0,Y_{1i}>0) \end{aligned}$$

As before, the left-hand side of (3) corresponds to $\mathrm{E}(Y_{1i} - Y_{0i})$ because of randomized treatment assignment. The expectation over (Y_{1i}, Y_{0i}) is with respect to the four events NP, S_1, S_2 and P. The last term is the intensive margin effect [IME], the first two are the extensive margin effect [EME], though as noted most models used in the literature will eliminate one of these.

2.2 *Nonparametric identification*

A comparison between (1) and (3) shows that they are distinct decompositions even under the monotone treatment response assumption. To highlight the difference, section 3 applies (1) and (3) to some of the most common corner solution response models in the literature. In those models, the decomposition based on joint potential outcomes is identified because considerable structure is imposed through functional form restrictions. Alternative identifying assumptions are discussed in the concluding section. In this section, it is shown that the decomposition is not identified nonparametrically: Experimental data combined with a monotone treatment response assumption alone do not point-identify all the objects of interest involved in the decomposition. The considerations below use the weaker monotone treatment response assumption that the treatment effect is either positive for all switchers, or negative.

A more compact notation will facilitate the exposition. Define the population fractions of switchers $\pi^S \equiv \mathrm{Pr}(Y_{i0} = 0, Y_{i1} > 0)$ and participants $\pi^P \equiv \mathrm{Pr}(Y_{i0} > 0, Y_{i1} > 0)$. Similarly, define mean potential outcomes of switchers $(\bar{Y}_0^S, \bar{Y}_1^S)$ and of participants $(\bar{Y}_0^P, \bar{Y}_1^P)$ (for instance, $\bar{Y}_0^S = \mathrm{E}(Y_{i1}|Y_{i0} = 0, Y_{i1} > 0)$). Then, the decomposition of the average treatment effect might be written as

$$\mathrm{ATE} = \pi^S \bar{Y}_1^S + \pi^P(\bar{Y}_1^P - \bar{Y}_0^P) = \pi^S \mathrm{ATE}^S + \pi^P \mathrm{ATE}^P \tag{4}$$

for the case $\bar{Y}_0^S = 0$, i.e. the case of group S_2 having mass zero. The reverse case $(\bar{Y}_0^S > 0, \bar{Y}_1^S = 0)$, i.e. group S_1 having mass zero, will not be considered — being symmetric, it gives no additional insights. Given the monotone treatment response assumption that one of the two cases holds, population regression of $D_i \equiv \mathbb{1}(Y_i > 0)$ on T_i reveals which it is: The difference $\mathrm{E}(D_i|T_i = 1) - \mathrm{E}(D_i|T_i = 0)$ corresponds to the difference between the S_1-fraction and the S_2-fraction in the population. Thus, if $\mathrm{E}(D_i|T_i = 1) > \mathrm{E}(D_i|T_i = 0)$, it must be that the S_2-fraction has mass zero.

Population regression of D_i on T_i can then be used to determine the population fractions of switchers and participants

$$\pi^S = \mathrm{E}(D_i|T_i = 1) - \mathrm{E}(D_i|T_i = 0) \quad \text{and} \quad \pi^P = \mathrm{E}(D_i|T_i = 1)$$

The term $\bar{Y}_0^P$ is also identified by the data; $\bar{Y}_0^P = \mathrm{E}(Y_i|D_i = 1, T_i = 0)$. The problem is identification of $\bar{Y}_1^S$ and $\bar{Y}_1^P$ for which only one quantity exists in the data, $\mathrm{E}(Y_i|D_i = 1, T_i = 1)$:

$$\mathrm{E}(Y_i|D_i = 1, T_i = 1) = \omega^S \bar{Y}_1^S + (1 - \omega^S)\bar{Y}_1^P, \quad \omega^S = \frac{\pi^S}{\pi^S + \pi^P}$$

Thus, it is impossible to disentangle them without making more assumptions; it follows that the decomposition is not identified nonparametrically — it is not possible to attribute a fraction of ATE to the extensive or intensive margin. However, since ATE, π^S and π^P are all identified, it is possible to derive sharp bounds for ATE^S and ATE^P using (4).

The bounds depend on the sign of ATE. Assume ATE < 0 first. Since $\mathrm{ATE}^S > 0$ (because S_2 is ruled out, as before), this means ATE^P must be negative. The domain of ATE^S is the positive real line $(0; \infty)$; the domain of ATE^P is the interval $(-\bar{Y}_0^P; 0)$.

Substituting the limits of these intervals into (4) the identification regions for the objects of interest reduce to

$$\text{ATE}^S \in \left(0, (\text{ATE} + \pi^P \bar{Y}_0^P)/\pi^S\right), \quad \text{ATE}^P \in \left(0, \text{ATE}/\pi^P\right)$$

The bounded regions are strictly smaller than the supports, and therefore informative.

Consider now ATE > 0. The data do not reveal the sign of the average treatment effect for participants. A strong monotone treatment response assumption would restrict it to be positive, so that ATE^P's domain would be $(0, \infty)$. In that case, a similar argument to the one above gives the following intervals for the ATE in the two population groups:

$$\text{ATE}^S \in \left(0, \text{ATE}/\pi^S\right), \quad \text{ATE}^P \in \left(0, \text{ATE}/\pi^P\right)$$

If one is unwilling to make this strong monotonicity assumption, the possibility of a negative ATE for participants has to be taken into account, which widens the domain of ATE^P; identification intervals are also widened in consequence but remain informative. If ATE^P < 0 this bounding strategy does not reduce ATE^P's domain which remains $(-\bar{Y}_0^P, 0)$. Combined with the previous result it follows that ATE$^P \in (-\bar{Y}_0^P, \text{ATE}/\pi^P)$, while ATE$^S \in (0, (\text{ATE} + \pi^P \bar{Y}_0^P)/\pi^S)$.

As in other partial identification settings (Manski, 2003), the availability of an additional discrete exogenous variable, say X_i, could tighten the upper bound for ATE^S further if X_i was related to the probability of being a switcher. Bounds for the average treatment effect could then be calculated for every value of the exogenous variable. The new upper bound for ATE^S resulting from adding the conditional-on-X_i bounds weighted by the mass points of X_i could be smaller than the ones given above. Similar arguments can be made for the bounds of participants.

3 Decomposing ATE in some common structural models

This section illustrates the decomposition based on joint potential outcomes for a class of models in which the objects of interest are point-identified.

3.1 *Tobit model*

The Tobit model (Tobin, 1958) is arguably the most popular model for corner solution dependent variables. It consists of three parts: a latent variable with a linear index structure, a distributional assumption for the error and an observation mechanism. These are

$$Y_i^* = \beta_0 + \beta_1 T_i + U_i \ , \quad U_i|, T_i \sim N(0, \sigma^2) \ , \quad Y_i = \max(0, Y_i^*) \tag{5}$$

Consider the case $\beta_1 > 0$, which imposes that Y_i is non-decreasing in T_i. The ATE in this model is

$$\mathrm{E}(Y_i|T_i = 1) - \mathrm{E}(Y_i|T_i = 0) = \Phi_1 \left(\beta_0 + \beta_1 + \sigma\phi_1/\Phi_1\right) - \Phi_0 \left(\beta_0 + \sigma\phi_0/\Phi_0\right)$$

where Φ_1, Φ_0 are the cdf of the standard normal distribution evaluated at $(\beta_0 + \beta_1)/\sigma$ and β_0/σ, respectively, and ϕ_1, ϕ_0 are the corresponding pdf. The conventional decomposition (1) would split this between extensive and intensive margin as follows

$$\widetilde{\mathrm{IME}} = (\beta_1 + \sigma\phi_1/\Phi_1 - \sigma\phi_0/\Phi_0)\Phi_0 \quad \widetilde{\mathrm{EME}} = (\Phi_1 - \Phi_0)(\beta_0 + \beta_1 + \sigma\phi_1/\Phi_1)$$

As was discussed before, it is difficult to assign a causal interpretation to $\widetilde{\mathrm{IME}}$ and $\widetilde{\mathrm{EME}}$. In the Tobit model, the distribution of potential outcomes of an individual is completely determined by her realization of the stochastic part U_i. Assume $\beta_1 > 0$, for concreteness. Then, individuals with U_i smaller than $-\beta_0 - \beta_1$ never have a positive Y_i; they constitute group NP (see Fig. 1). Similarly, if U_i lies between $(-\beta_0 - \beta_1)$ and $(-\beta_1)$, individuals are group S_1 switchers. (S_2 switchers, i.e. individuals dropping out of participation because of treatment, are incompatible with the structure of the model when $\beta_1 > 0$.) $\widetilde{\text{EME}}$ correctly identifies the fraction of switchers $(\Phi_1 - \Phi_0)$ but fails to attribute the correct ATE. Rather, it assigns to them the average Y_i in the population of switchers and participants. This overestimates their contribution, as switchers' U_i are in the bottom tail of the error distribution among those with $Y_{1i} > 0$. Their true ATE is $\beta_0 + \beta_1 + \text{E}(U_i | -\beta_0 - \beta_1 \leq U_i < \beta_1)$. Thus the correction term is the expectation of a doubly-truncated normal variable. Table 1 contains the features of switchers and participants in the Tobit model. Multiplying the second by the fourth row gives the causal IME and EME.

- - - Figure 1 and Table 1 about here - - -

Consider a numerical example to illustrate the difference which using the decomposition based on joint potential outcomes can make. Suppose the DGP is (5) with $\beta_0 = 0, \beta_1 = 1, \sigma^2 = 1$. Then the ATE is about 0.68. The conventional decomposition assigns about 0.24 to the intensive and 0.44 to extensive margin effect. In contrast, the decomposition into causally meaningful margins reveals that of the total ATE of 0.68, 0.5 is due to the intensive and only 0.18 due to the extensive margin effect. The intensive margin contribution, which was only 36% using the old decomposition, is thus really 73%.

Similarly stark discrepancies are possible in practice. For instance, McDonald and Moffitt's (1980) application examined the effect of a negative income tax on working hours reductions by estimating a Tobit model. Using their decomposition, it assigned 22% of the estimated reduction in working hours to the extensive margin. A follow-up article by Moffitt (1982) reevaluated the same data. In this article, he modified the Tobit model to account for a model of labor market frictions. Incidentally, this leads to the same formulas for the decomposition as the ones using the decomposition based on joint outcomes presented in Table 1. Applying this decomposition, he now concluded that the extensive margin was responsible not for 22%, but for 84% of the reduction. The present article shows that even in the absence or misspecification of the specific labor market frictions model postulated in Moffitt (1982), the causal extensive margin contribution is 84%.

Coming back to the numerical illustration from before, the example DGP can also be used to illustrate the bounds discussed in the previous section. Here, the average treatment effect for participants, ATE^P, is 1 ($= \beta_1$), and the ATE for switchers, ATE^S, is about 0.54. A researcher ignoring the DGP and reluctant to make any assumptions on it can conclude that $\text{ATE}^P \in (0; 1.36)$ and $\text{ATE}^S \in (0; 2)$.

3.2 *Selection and two-part models*

There are several generalizations and alternatives to the Tobit model that are formulated in the same framework (Cragg, 1971; Heckman, 1979; Duan et al., 1983; Amemiya, 1985). One variant postulates

$$\begin{aligned} Y_i &= D_i \exp(\beta_0 + \beta_1 T_i + U_i) \\ D_i &= \mathbb{1}(\alpha_0 + \alpha_1 T_i + V_i) \end{aligned} \tag{6}$$

with $(U_i, V_i) \sim BVN(0, 0, \sigma^2, 1, \rho)$. The exponential transformation of the right-hand side of Y_i ensures positivity of Y_i. It is common to refer to model (6) as 'selection model' when the errors are correlated, and as 'two-part' model when errors are independent (Hay and Olsen, 1984). It is clear from (6) that here, in contrast to the Tobit model, the signs of extensive and intensive margin need not be the same, as they are driven by the signs of α_1 and β_1, respectively. Moreover, (6) models the participation decision (the equation for D_i) and the outcome conditional on participation as (potentially) driven by two separate errors (U_i and V_i). Analogously to the Tobit model, the population fraction of groups are $(\Phi_1 - \Phi_0)$ for switchers and Φ_0 for participants, where now $\Phi_1 = \Phi(\alpha_0 + \alpha_1)$ and $\Phi_0 = \Phi(\alpha_0)$.

Consider the two-part model first, i.e. assume correlation $\rho = 0$. The essential feature of the two-part model is that because the errors are independent, the conditional error expectation $\text{E}(\exp(U_i)|V_i) = \exp(0.5\sigma^2)$ is the same for switchers and participants. This means that both decompositions coincide: A switcher has an expected treatment effect of $\exp(\beta_0 + \beta_1 + 0.5\sigma^2)$ which is just $\text{E}(Y_i|Y_i > 0, T_i = 1)$, and participants experience a percental change of $\exp(\beta_1) - 1$. Or, in terms of the analysis in (2), the selection bias in the COP effect vanishes in this model because $\mathrm{E}(Y_{0i}|Y_{1i} > 0) = \mathrm{E}(Y_{0i}|Y_{0i} > 0) = \exp(\beta_0 + 0.5\sigma^2)$.

The assumption of zero selection bias is unwarranted in most applications. While randomization prevents dependence between treatment and the errors, there is no experiment which could possibly break the potential dependence between U_i and V_i — and applications where the researcher can be certain that this dependence is absent seem difficult to envision.

Thus, consider the selection model which allows correlation between the errors. Since the model estimated in the application in the following section is a selection model with covariates, model (6) is rewritten to accommodate this feature. With covariates, model (6) is

$$Y_i = D_i \times \exp(X_i\beta + \beta_T T_i + U_i) \tag{7}$$

$$D_i = \mathbb{1}(Z_i\alpha + \alpha_T T_i + V_i \geq 0) \tag{8}$$

with $(U_i, V_i) \mid T_i, X_i, Z_i \sim BVN(0, 0, \sigma^2, 1, \rho)$. This distributional assumption implies that treatment and regressors are independent of the errors. Regressors are collected in two vectors Z_i and X_i with corresponding coefficient vectors α and β. No exclusion restriction is placed on covariates. In principle, they can be identical, disjoint or overlapping, although economic considerations will commonly lead to a set of overlapping, if not identical, variables.

The normality assumption implies a probit model for the decision to participate, $\Pr(D_i = 1|Z_i) = \Phi(Z_i\alpha + \alpha_T T_i)$. Then, for given Z_i, switchers are defined by values of V_i lying in the interval $S \equiv [-Z_i\alpha - \alpha_T; -Z_i\alpha)$, and participants by $V_i \in P \equiv [-Z_i\alpha; \infty)$. For observations with characteristics Z_i, the fractions of participants and that of switchers are

$$\Pr(V_i \in P) = \Phi(Z_i\alpha), \quad \Pr(V_i \in S) = \Phi(Z_i\alpha + \alpha_T) - \Phi(Z_i\alpha)$$

As seen previously, both decompositions assign the same value to the population fraction, but they differ in assigning average treatment effects for switchers and participants. For switchers, ATE^S conditional on covariates is

$$\text{ATE}^S = \text{ATE}(X_i, Z_i, V_i \in S) = \\ \exp(X_i\beta + \beta_T + 0.5\sigma^2)\frac{\Phi(\sigma\rho + Z_i\alpha + \alpha_T) - \Phi(\sigma\rho + Z_i\alpha)}{\Phi(Z_i\alpha + \alpha_T) - \Phi(Z_i\alpha)} \tag{9}$$

the correction term is the doubly-truncated expectation $\text{E}(\exp(U_i)|V_i \in S)$. Instead, the standard decomposition uses the expectation of Y_i given $D_i = 1$ and $T_i = 1$ for ATE^S. The expectation of Y_i conditional on participation is (cf. Terza, 1998)

$$\text{E}(Y_i|D_i = 1, Z_i, X_i) = \exp(X_i\beta + \beta_T T_i + 0.5\sigma^2)\frac{\Phi(\sigma\rho + Z_i\alpha + \alpha_T T_i)}{\Phi(Z_i\alpha + \alpha_T T_i)} \tag{10}$$

where the correction term is the simple truncated expectation $\text{E}(\exp(U_i)|X_i, V_i > -Z_i\alpha - \alpha_T T_i)$. Essentially, this produces the same pattern of discrepancies between decompositions as in the Tobit model, although $|\rho| < 1$ will lessen the magnitude of the difference (with the two-part model being the limit case). In this model, the relative size of the conventional extensive margin effect ($\widetilde{\text{EME}}$) relative to the causal one (EME) depends solely on the linear index of the participation equation $Z_i\alpha$, α_T and $\sigma\rho$, but not on β and β_T:

$$\frac{\widetilde{\text{EME}}}{\text{EME}} = \left(1 - \frac{\Phi(Z_i\alpha)}{\Phi(Z_i\alpha + \alpha_T)}\right) \Bigg/ \left(1 - \frac{\Phi(\sigma\rho + Z_i\alpha)}{\Phi(\sigma\rho + Z_i\alpha + \alpha_T)}\right) \tag{11}$$

Thus, $\rho > 0$ ($\rho < 0$) implies that the conventional distribution overestimates (underestimates) the causal EM. Also, for a given value of ρ, the extent of the discrepancy is increasing in both $Z_i\alpha$ and α_T.

For participants, the conditional average treatment effect is

$$\text{ATE}^P = \text{ATE}(X_i, Z_i, V_i \in P) = (\exp(\beta_T) - 1)\exp(X_i\beta)\frac{\Phi(\sigma\rho + Z_i\alpha)}{\Phi(Z_i\alpha)} \tag{12}$$

while the conventional decomposition prescribes

$$\exp(X_i\beta + \beta_T + 0.5\sigma^2)\frac{\Phi(\sigma\rho + Z_i\alpha + \alpha_T)}{\Phi(Z_i\alpha + \alpha_T)}\Phi(Z_i\alpha) - \exp(X_i\beta + 0.5\sigma^2)\Phi(\sigma\rho + Z_i\alpha)$$

Unconditional ATE for switchers and participants can be obtained by taking expectations over the distribution of (Z_i, X_i), e.g. $\text{ATE}^S = \text{E}_{Z_i,X_i}[\text{ATE}(X_i, Z_i, V_i \in S)]$.

4 An application: The trade effect of reducing the number of bureaucratic firm-entry-regulation procedures

This section applies the new decomposition to an empirical trade model. Traditionally, the determinants of trade volumes were estimated in a single-equation, constant-elasticity gravity model (Santos Silva and Tenreyro, 2006; Feenstra, 2008). However, the large fraction of zeros in aggregated trade datasets spanning many countries has motivated a new strand of empirical literature which favors a two-equations model (Helpman, Melitz and Rubinstein, 2008). The first equation addresses the zeros directly by modeling trade participation. The second equation models trade flows conditional on participation. The trade volume equation is specified as a traditional gravity model. With these equations, explanatory variables can influence trade flows at two country margins, the extensive margin –the decision to trade– and the intensive margin –average trade flows of trading country-pairs. In this application, I will analyze the trade effect of a hypothetical policy intervention which would reduce the number of bureaucratic procedures needed to set up a business legally.

The empirical model is the generalized Tobit model (7)-(8). The indicator variable D_i declares the presence or absence of trade between a directed country-pair i, and the variable Y_i will denote its trade volume ($Y_i \geq 0$). The term "directed country-pair" means here that for every pair of countries there are two observations: the exports of the first to the second and vice versa. The vector of variables explaining the decision to trade are Z_i, the variables explaining the trade volume X_i, and unobserved variables (as well as pure randomness) in the participation and volume equations are V_i and U_i, respectively. The set of variables X_i and Z_i can contain distinct elements, in principle. Indeed, much of the theoretic motivation for the two-equations model comes from the idea that zero trade flows are due to the impossibility of overcoming fixed costs which are necessary to establish trade (Hallak, 2006). This suggests that Z_i contains "fixed costs" and X_i "variable costs" of trading. In practice, however, the case seems less clear-cut as at the aggregate country-level the variables observed are not the "costs" directly, but rather rough proxies for them, such as distance between capital cities, which makes it hard to distinguish between fixed and variable costs. For instance, firm entry regulation measures such as the number of procedures should primarily be a fixed cost and not affect variable trading costs (Helpman, Melitz and Rubinstein, 2008). But Djankov et al. (2002) relate such costs to corruption and shadow economies which are likely to affect variable trade costs as well. Baranga (2009, fn. 9) provides an alternative argument against excluding firm entry regulation variables from X_i: "[A] country with higher regulatory barriers may also be more likely to be a higher tax environment, which would be expected to reduce the profitability of exporting at the intensive margin too. Countries with more regulation might also be more likely to use quantitative trade restrictions such as import or export licenses, or other non-tariff barriers, which would also affect the intensive margin, but are typically not controlled for." Thus, no exclusion restrictions will be placed on the variables here, so that $X_i = Z_i$. Finally, adopting the assumption of bivariate normal errors facilitates comparison with previous studies.

Estimation of (7)-(8) can be carried out by full information ML. Here, I will estimate the model by the standard "Heckit" two-step procedure, which in a first step estimates (8) by Probit ML, and uses the estimated $\hat{\alpha}$ to estimate

$$\ln(Y_i) = X_i\beta + \sigma\rho\phi(X_i\hat{\alpha})/\Phi(X_i\hat{\alpha}) + \epsilon_i \tag{13}$$

by OLS in the sample with $D_i = 1$ (second step). The estimating equation (13) can be seen as an approximation to moment-based estimation using the condition $\mathrm{E}[Y_i - \mathrm{E}(Y_i|D_i = 1, X_i)|X_i] = 0$, where $\mathrm{E}(Y_i|D_i = 1, X_i)$ is given in (10). In particular, the inverse Mills ratio is a first order approximation to the multiplicative correction term in (10) (Greene, 1998).

The objects of interest are ATE, IME and EME associated with the policy intervention of reducing the number of bureaucratic procedures; they can be computed from the parameter estimates of the probit equation and of (13) using the formulas provided in the preceding section.

4.1 *Data*

The data is taken from the study by Helpman, Melitz and Rubinstein (2008), which the authors kindly make publicly available on the internet (the data can be downloaded from http://www.economics.harvard.edu/faculty/helpman/Data_Sets_Helpman). It is pooled from different sources, including Feenstra's World Trade Flows, the Penn World Tables and the World Bank's World Development Indicators; and is described in detail in Helpman, Melitz and Rubinstein's (2008) Appendix I. Part of their analysis uses country-level data on regulation costs of firm entry collected by Djankov et al. (2002). Specifically, they create two dummy variables indicating a country-pair having high regulation costs. The first is based on the number of bureaucratic procedures it takes to set up a business in a given country, and the number of days that it takes to complete these procedures. The variable equals one when both countries in the pair are above the median according to these criteria. Similarly, the second dummy equals one when importer and exporter are above the median according to regulation costs as measured as a percentage of countries' GDP. In addition to these two binary variables, I use the sum of the number of procedures required in the importing and in the exporting country of a pair; this is the variable of interest in this application.

- - - Table 2 about here - - -

Descriptive statistics for the variables used in this analysis are presented in Table 2. They correspond to the year 1986 (with the exception of the regulation cost variables, which are from 1999). The dataset consists of 11,978 country-pairs of 106 exporter countries and 114 importer countries. The asymmetry stems from countries serving the whole (sampled) world as exporters. As it is necessary to estimate sets of exporter and importer fixed effects to control for multilateral effects (Anderson and van Wincoop, 2003; Feenstra; 2004), all-world exporters were dropped from the dataset to avoid perfect prediction in the decision-to-trade equation. As can be seen from the number of observations for the logarithm of bilateral trade, only 6,572 out of 11,978 (or 55%) of the country-pairs engage in trade. To explain trade flows, I broadly follow the specification of Helpman, Melitz and Rubinstein. The regressors include great-circle distance between capitals in log-Kilometers (*Distance*), the gravity equation variable par excellence. To capture geography-related trade costs further, the indicator variables *Landlock* (at least one country in pair is landlocked), *Island* (at least one country in pair is an island), and *Land border* (countries share a common border) are used in the specification. Cultural and historical similarities are proxied by the dummy variables *Legal* (origin of legal systems of the countries are the same), *Language* (countries have common language), *Colonial ties* (one country was/is the other's colony) and *Religion*, a continuous index ranging from 0 to 1, which aggregates the similarity in the composition of Catholics, Protestants and Muslims in the countries. As discussed above, regulation costs are mapped by the indicators *Reg. costs (% GDP)* and *Reg. costs (days & proc.)*, as well as *No. of procedures*.

4.2 *Estimation results*

The estimated coefficients of a two-stage Heckit procedure are reported in Table 3. The explanatory variables included a set of importer and exporter fixed effects. Due to collinearity, it was not possible to estimate a separate exporter fixed effect for Chad in neither of the two equations. Thus, there is only one joint exporter fixed effect for South Africa, the base-category country, and Chad.

- - - Table 3 about here - - -

Despite the slightly different data set and specification, the coefficients in Table 3 are very similar to the results of Helpman, Melitz and Rubinstein (2008). Specifically, I would like to focus on the effect of the following policy intervention: cutting back two bureaucratic procedures. Two procedures correspond to about half a standard deviation of the variable, and one can think of the intervention as both importer and exporter country eliminating each one bureaucratic hurdle. The effect of the number of procedures on trade is large, judging from the results in Table 3: Two procedures less is expected to increase the probability of trading by about 12%-points for the average country-pair (here: $X_i\hat{\alpha} = 0.46$); a trading country-pair is expected to increase its trade volume by about 146% ($= \exp(-2\times -0.45)-1$) on average in response to such a policy intervention. Of course, such causal interpretations are only valid if all model assumptions hold; in particular, if regressors are independent of errors.

Consider as an example country-pairs for which the average predicted probability of participation is 0.5. Observations with such an average probability include trade from Romania to Bolivia or from Ethiopia to South Korea, which is positive; as well as trade from Romania to Honduras or from Cambodia to South Korea, which is zero. Table 4 shows the estimated ATE, EME and IME for such country-pairs. In the first row of Table 4 it is assumed that $\sigma\rho = 0$. As discussed previously, the two decompositions (1) and (3) of the ATE coincide in this case. The model has been estimated using $\ln(y)$ as the dependent variable. To retransform predictions to levels, an estimate of σ^2 is needed. Such an estimate is not directly available because the two-step estimator only gives $\widehat{\sigma\rho}$. Therefore, Duan's (1983) smearing estimator was used to obtain predictions in levels. The total effect of 893 corresponds to an increase in expected trade flows of about 140%. Of the 893, the extensive margin contributes 39%.

But assume first that $\sigma\rho = 0$, and calculate the total effect and its decomposition under this premise (first row in Table 4). As discussed previously, the two decompositions (1) and (3) of the ATE coincide in this case. The model has been estimated using $\ln(y)$ as the dependent variable. To retransform predictions to levels, an estimate of σ^2 is needed. Such an estimate is not directly available because the two-step estimator only gives $\widehat{\sigma\rho}$. Therefore, Duan's (1983) smearing estimator was used to obtain predictions in levels. The total effect of 893 corresponds to an increase in expected trade flows of about 140%. Of the 893, the extensive margin contributes 39%.

- - - Table 4 about here - - -

The estimated coefficient on the inverse Mills ratio, $\widehat{\sigma\rho}$, is 0.20 with a p-value of 2.4%, suggesting that there is some dependence between errors. Rows (2)-(5) of Table 4 show estimates of the total effect and its decomposition with correlation. Row (2) contains an estimate as may be found in previous literature. It uses the standard decomposition (1), and approximates the conditional expectation function $\mathrm{E}(Y_i|D_i = 1, X_i)$ by $\hat{\eta}\exp(X_i\hat{\beta} + \widehat{\sigma\rho}\hat{\lambda}_i)$, where $\hat{\eta}$ is used to denote the smearing estimate of $\mathrm{E}(\exp(\epsilon)|x)$ and $\hat{\lambda}_i$ is the estimated inverse Mills ratio for X_i. Row (4) contains results of the same decomposition but using the exact functional forms of section 3.2. It turns out that the difference between using approximate (with inverse Mills ratio) and exact expectations (with multiplicative correction term) is negligible in this application.

Regarding the relative contribution of extensive and intensive country margin to the total effect, the differences between omitting correlation and taking it into account are small in this case (39% vs. 40%). The magnitude of the total effect is also quite close to the one ignoring correlation. Thus, the relative effect necessarily needs to be smaller, since the expectations conditioning on the error correlation are larger than the ones omitting the adjustment factor because $\Phi(\sigma\rho + X_i\alpha)/\Phi(X_i\alpha) > 1$ for positive ρ. The difference is almost 20%-points (139% vs. 122%).

Comparing the results of the conventional decomposition to the one proposed in this paper based on country-types, the shift in the contribution of the margins is clearly visible. The extensive margin contribution decreases by 6%-points going from row (4) to (5), a 15% difference. The reason for this drop is that with positive correlation, the doubly-truncated expectation of the error underlying the new decomposition will always be lower than the error expectation truncated from below at the upper bound of the doubly-truncated expectation which underlies the conventional method. The total effect is the same. While for the approximations in rows (2) and (3) this identity is slightly veiled, it holds precisely when using the exact functional forms in rows (4) and (5).

The overestimation of the extensive margin is not limited to this group of country-pairs, of course. For instance, using country-pairs at the mean $X_i\hat{\alpha}$, the overestimation of the extensive margin by the standard decomposition is over 20%. Using the estimated value of ρ $(= 0.2)$, and of α_T $(= -2 \times -0.2 = 0.4)$, Fig. 2 plots the ratio of conventional versus causal EME (cf. Eq. 11) over some of the range of the participation probability, $\Phi(X_i\alpha)$. The increase in overestimation is quite steep.

- - - Figure 2 about here - - -

5 Discussion

This paper presented a decomposition of average treatment effects in corner solution models into extensive and intensive margins based on the joint distribution of potential outcomes. The new decomposition is a weighted sum of the ATE of subgroups of the population —switchers and participants—, and it differs markedly from the traditional decomposition, which lacks an interesting causal interpretation. This was demonstrated in a numerical example for the Tobit model, and in a substantive application to international trade flows for a generalization of the Tobit model. By relying on very strong distributional assumptions, these models display tractable closed forms which are useful for both illustration and for comparison with previous research.

However, the decomposition of treatment effects presented here is also applicable to semiparametric models. One such class are latent factor structure models (Aakvik, Heckman and Vytlacil, 2005). These models relax the functional form assumptions, but in turn require an exclusion restriction. I.e., an instrumental variable is needed which affects switchers but not participants. An alternative assumption which would have identifying power within linear-index models would be that participants and switchers display (different) index heteroskedasticity, as in Klein and Vella (2009). Both functional form and exclusion restriction assumptions are non-refutable, but their plausibility might differ depending on the application.

While the choice of decomposition matters in general, it makes no difference under the two-part model. The reason for this is the assumed error independence in the class of two-part models I considered. This assumption is unattractive, but it seems that it is not a necessary ingredient of two-part models. Duan et al. (1984) provide an example of a two-part model with correlated errors where the correlation parameter does not enter the likelihood function. Thus, consistent parameter estimates can still be obtained conveniently by separate probit and linear regressions. In such two-part models, the correct decomposition *would* differ from the traditional — however, neither decomposition could be calculated (nor even the ATE) as the correlation parameter is unavailable. Thus, this hardly seems to make the two-part model more attractive for causal inference.

The causally meaningful decomposition of the ATE does not come free of cost: It requires more assumptions than needed for the ATE alone, as it deals with potential outcomes jointly. As applied work often makes assumptions that go well beyond the required for the decomposition, however, choosing the correct decomposition does seem to be almost free of cost. For instance, all the applied articles cited in the introduction imposed enough structure to point-identify the causal decomposition.

Finally, while this article examined the decomposition into effects at margins in a simple experimental setting where nonparametric identification fails, other experimental settings can be devised which have point-identifying power under weaker conditions. Pre-treatment measurements of Y_i is one such setting. If it can be ensured that individual-specific unobservables do not vary over time ($U_i = U_{it}$ for periods $t = 0, 1$), the decomposition is nonparametrically identified.

References

Aakvik, Arild, James J. Heckman and Edward J. Vytlacil (2005), "Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs", *Journal of Econometrics*, **125**(1), pp. 15-51.

Amemiya, Takeshi (1985), *Advanced Econometrics*, Harvard University Press.

Anderson, James E. and Eric van Wincoop (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review*, **93**(1), pp. 170-192.

Angrist, Joshua D. (2001), "Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice", *Journal of Business and Economic Statistics* **19**(1), pp. 2-16.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009), *Mostly Harmless Econometrics*, Princeton University Press.

Variables:

Baranga, Thomas (2009), "Unreported Trade Flows and Gravity Equation Estimation", unpublished manuscript, IRPS, UC San Diego.

Cameron, A. Colin and Pravin K. Trivedi (2005), *Microeconometrics*, Cambridge University Press.

Caspi, Avshalom, Bradley R. Entner Wright, Terrie E. Moffitt, Phil A. Silva (1998), "Early Failure in the Labor Market: Childhood and Adolescent Predictors of Unemployment in the Transition to Adulthood" *American Sociological Review* **63**(3), pp. 424-451.

Chen, Songnian (2010), "Non-Parametric Identification and Estimation of Truncated Regression Models", *Review of Economic Studies*, **77**(1), pp. 127-153.

Chiburis, Richard C. (2010), "Semiparametric bounds on treatment effects", *Journal of Econometrics*, **159**(2), pp. 267-275.

Co, Catherine Yap (2010), "Intra- and inter-firm US trade ", *International Review of Economics and Finance*, **19**(2), pp. 260-277.

Cragg, John G. (1971), "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods", *Econometrica*, **39**(5), pp. 829-844.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes and Andrei Shleifer, (2002), "The Regulation of Entry", *Quarterly Journal of Economics*, **117**(1), pp. 1-37.

Dow, William H. and Edward C. Norton (2003), "Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions", *Health Services and Outcomes Research Methodology*, **4**(1), pp. 5-18.

Duan, Naihua (1983), "Smearing estimate: a nonparametric retransformation method", *Journal of the American Statistical Association*, **78**(3), pp. 605-610.

Duan, Naihua, Willard G. Manning, Jr., Carl N. Morris, Joseph P. Newhouse (1983), "A Comparison of Alternative Models for the Demand for Medical Care", *Journal of Business and Economic Statistics*, **1**(2), pp. 115-126.

Duan, Naihua, Willard G. Manning, Jr., Carl N. Morris, Joseph P. Newhouse (1984), "Choosing between the Sample-Selection Model and the Multi-Part Model", *Journal of Business and Economic Statistics*, **2**(3), pp. 283-289.

Engelhardt and Kumar (2007), "Employer matching and 401(k) saving: Evidence from the health and retirement study", *Journal of Public Economics*, **91**(10), pp. 1920-1943.

Fan, Yanqin and Jisong Wu (2010), "Partial Identification of the Distribution of Treatment Effects in Switching Regime Models and its Confidence Sets", *Review of Economic Studies*, **77**(3), pp. 1002-1041.

Feenstra, Robert C. (2004), *Advanced International Trade: Theory and Evidence*, Princeton University Press.

Feenstra, Robert C. (2008), "Gravity Equation", in: *The New Palgrave Dictionary of Economics*, Second Edition, Eds. Steven N. Durlauf and Lawrence E. Blume, Palgrave Macmillan.

Felbermayr, Gabriel J. and Wilhelm Kohler (2006), "Exploring the intensive and extensive margin of world trade", *Review of World Economics 2006*, **142**(4), pp. 642-674.

Greene, William H. (1998), "Sample selection in credit-scoring models", *Japan and the World Economy*, **10**(3), pp. 299-316.

Greene, William H. (2008), *Econometric Analysis*, 6th edition, Prentice Hall.

Hallak, Juan Carlos (2006), "Product quality and the direction of trade", *Journal of International Economics*, **68**(1), pp. 238-365.

Hastings, Justine and Ebonya Washington (2010), "The First of the Month Effect: Consumer Behavior and Store Responses." *American Economic Journal: Economic Policy*, **2**(2), pp. 142-162.

Hay, Joel W. and Randall J. Olsen (1984), "Let Them Eat Cake: A Note on Comparing Alternative Models of the Demand for Medical Care", *Journal of Business and Economic Statistics*, **2**(3), pp.279-282.

Heckman, James J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, **47**(1), pp. 153-161.

Helpman, Elhanan, Marc Melitz and Yona Rubinstein (2008), "Estimating Trade Flows: Trading Partners and Trading Volumes", *Quarterly Journal of Economics*, **123**(2), pp. 441-487.

Joulfaian, David (2000), "Corporate Income Tax Evasion and Managerial Preferences", *Review of Economics and Statistics*, **82**(4), pp. 698-701.

Kenkel, Donald S. (1991), "Health Behavior, Health Knowledge, and Schooling", *Journal of Political Economy*, **99**(2), pp. 287-305.

Klein, Roger and Francis Vella (2009), "A Semiparametric Model for Binary Response and Continuous Outcomes Under Index Heteroscedasticity", *Journal of Applied Econometrics*, **24**(5), pp. 735-762.

Liu, Xuepeng (2009), "GATT/WTO Promotes Trade Strongly: Sample Selection and Model Specification", *Review of International Economics*, **17**(3), pp. 428-446.

Manski, Charles F. (1997), "Monotone treatment response", *Econometrica*, **65**(6), pp. 1311-1334.

Manski, Charles F. (2003), *Identification for Prediction and Decision*, Harvard University Press.

McDonald, John F. and Moffitt, Robert A. (1980), "The Uses of Tobit Analysis", *Review of Economics and Statistics*, **62**(2), pp. 318-321.

Moffitt, Robert A. (1982), "The Tobit Model, Hours of Work and Institutional Constraints", *Review of Economics and Statistics*, **64**(3), pp. 510-515.

Santos Silva, João M.C. and Silvana Tenreyro (2006), "The Log of Gravity", *Review of Economics and Statistics*, **88**(4), pp. 641-658.

Sousa-Poza, Alfonso and Alexandre Ziegler (2003), "Asymmetric information about workers' productivity as a cause for inefficient long working hours", *Labor Economics*, **10**(6), pp. 727-747.

Terza, Joseph V. (1998), Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects, *Journal of Econometrics*, **84**(1), pp. 129-154.

Tobin, James (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, **26**(1), pp. 24-36.

Trejo, Stephen S. (1993), "Overtime Pay, Overtime Hours, and Labor Unions", *Journal of Labor Economics*, **11**(2), pp. 253-278.

Vytlacil, Edward (2002), "Independence, Monotonicity and Latent Index Models: An Equivalence Result", *Econometrica*, **77**(1), pp. 331-341.

Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

Table 1: Features of participants and switchers in the Tobit model

	Participants ($Y_{0i} > 0, Y_{1i} > 0$)	Switchers ($Y_{0i} = 0, Y_{1i} > 0$)
U_i	$(-\beta_0, \infty)$	$(-\beta_0 - \beta_1, -\beta_0)$
$\Pr(Y_{1i}, Y_{0i})$	Φ_0	$\Phi_1 - \Phi_0$
$Y_{1i} - Y_{0i}$	β_1	$\beta_0 + \beta_1 + U_i$
$\mathrm{E}(Y_{1i} - Y_{0i})$	β_1	$\beta_0 + \beta_1 + \sigma \frac{\phi_1 - \phi_0}{\Phi_1 - \Phi_0}$

Notes: $\Phi_T = \Phi(\beta_0 + \beta_1 T)$, $\phi_T = \phi(\beta_0 + \beta_1 T)$, for $T = 0, 1$. $\Phi(\cdot)$ is the standard normal cdf, $\phi(\cdot)$ the standard normal pdf. The Tobit model in this table has the latent variable $Y_i^* = \beta_0 + \beta_1 T_i + U_i$, with $U_i | T_i \sim N(0, \sigma^2)$ and $\beta_1 > 0$.

Table 2: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	No. of obs.
Bilateral trade	79,804.40	991,081.28	0	74,558,336	11,978
Log of Bilateral Trade	8.33	3.04	1.61	18.13	6,572
Distance	4.17	0.8	0.3	5.66	11,978
Land border	0.02	0.15	0	1	11,978
Island	0.17	0.37	0	1	11,978
Landlock	0.36	0.48	0	1	11,978
Legal	0.36	0.48	0	1	11,978
Language	0.26	0.44	0	1	11,978
Colonial ties	0.01	0.09	0	1	11,978
Religion	0.17	0.25	0	0.99	11,978
No. of procedures	19.59	4.86	4	36	11,978
Reg. costs high (%GDP)	0.33	0.47	0	1	11,978
Reg. costs high (days & proc.)	0.12	0.33	0	1	11,978

Source: Data are from Helpman, Melitz and Rubinstein (2008), available online. See text Section 4.1.

Table 3: Estimated coefficients — Two-equations model of bilateral trade

Regression	$\Pr(d = 1 \mid x)$	$\mathrm{E}(\ln y \mid y > 0, x)$
Method	ML (Probit)	OLS
	(1)	(2)
Distance	-0.62**	-1.22**
	(0.03)	(0.04)
Land border	-0.16	0.67**
	(0.13)	(0.14)
Island	-0.54*	-0.44
	(0.24)	(0.25)
Landlock	-0.14	-0.42*
	(0.12)	(0.17)
Legal	0.15**	0.55**
	(0.04)	(0.06)
Language	0.32**	0.19**
	(0.06)	(0.07)
Colonial ties	-0.02	0.89**
	(0.35)	(0.19)
Religion	0.33**	0.32**
	(0.09)	(0.11)
No. of procedures	-0.20**	-0.45**
	(0.03)	(0.05)
Reg. costs high (%GDP)	-0.27**	-0.13
	(0.08)	(0.09)
Reg. costs high (days & proc.)	-0.16*	-0.25*
	(0.07)	(0.11)
Inv. Mills ratio		0.20*
		(0.09)
R^2	0.57	0.69
$\log L$	-3,580	-12,760
Observations	11,978	6,572

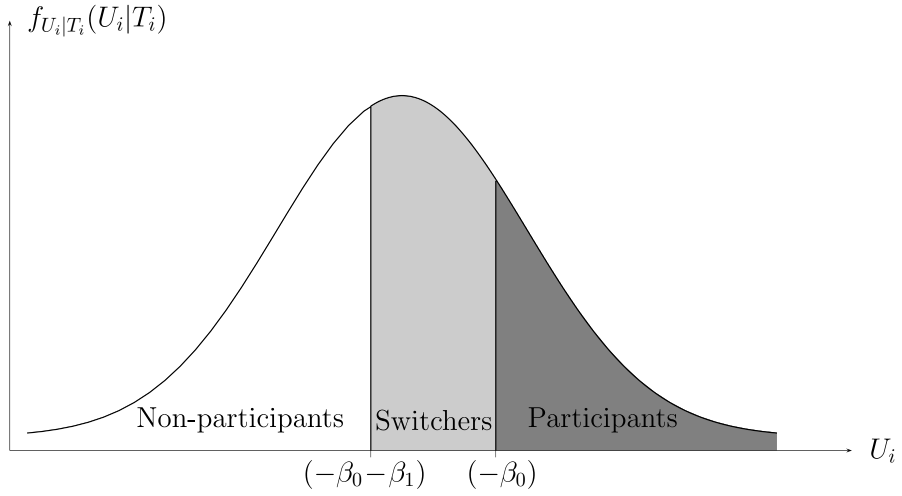
Notes: Robust standard errors in parentheses. $*$ and $**$ denote statistical significance on the 5% and 1% level. Additional regressors include a constant term and a complete set of importer fixed effects and of exporter fixed effects.

Table 4: Total trade effects and decomposition into country margins

Decomposition	TE	TE/E(y|x)	EME (% of TE)	IME (% of TE)
(1) No error correlation	893	139%	347	546
			(39%)	(61%)
(2) Conventional decomposition, approx.	888	122%	353	535
			(40%)	(60%)
(3) Decomposition by country-type, approx.	892	123%	308	584
			(35%)	(65%)
(4) Conventional decomposition, exact	879	122%	349	530
			(40%)	(60%)
(5) Decomposition by country-type, exact	879	122%	302	577
			(34%)	(66%)

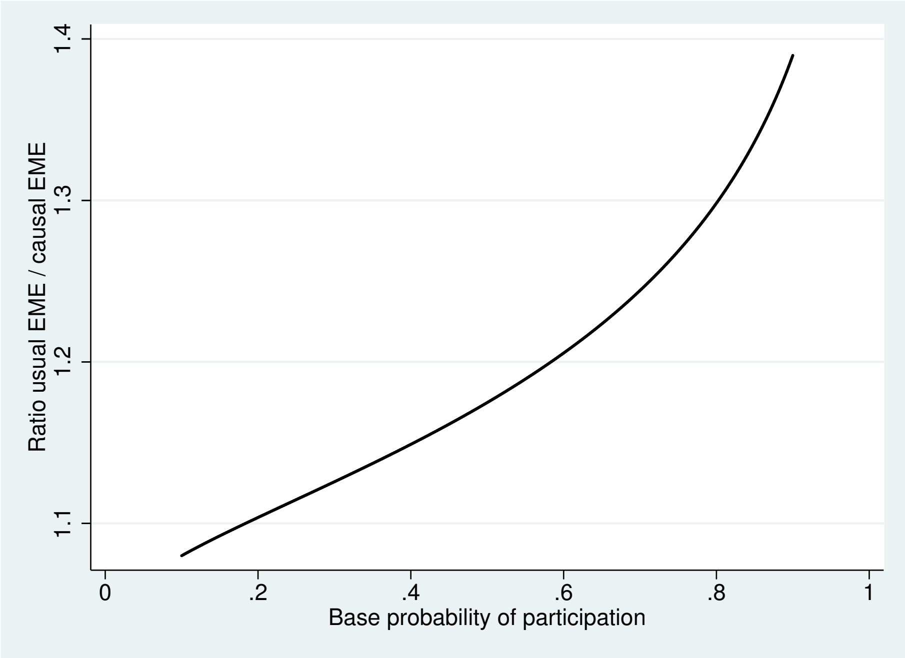
Notes: Own calculations based on results from Table 3. TE stands for Total Effect, EME for Extensive Margin Effect, and IME for Intensive Margin Effect. Formulas are discussed in Section 3.2. Effects are for a country-pair with $X\hat{\alpha} = -0.2$ and $X\hat{\beta} = 5.32$.

Figure 1: Population groups by U_i in the Tobit model



Notes: The Tobit model in this figure has the latent variable $Y_i^* = \beta_0 + \beta_1 T_i + U_i$, with $U_i|T_i \sim N(0, \sigma^2)$ and $\beta_1 > 0$.

Figure 2: Overestimation of extensive margin effect (EME) in estimated trade model



Notes: Own calculations based on results from Table 3. "Base probability of participation" means participation probability without treatment (reduction in "No. of proc."). Base probability of participation plotted over range 0.1–0.9. "Ratio usual EME / causal EME" is $\widetilde{\text{EME}}/\text{EME}$ as in Eq. (11), calculated for estimated values of $\widehat{\sigma\rho}$ (= 0.2) and $\hat{\alpha}_T$ ($= -2 \times -0.2 = 0.4$).

Working Papers of the Socioeconomic Institute at the University of Zurich

The Working Papers of the Socioeconomic Institute can be downloaded from http://www.soi.uzh.ch/research/wp_en.html

1013	The Trade Effects of Endogenous Preferential Trade Agreements, Peter Egger, Maric Larch, Kevin E. Staub, Rainer Winkelmann, November 2010, 48 p.
1012	A Causal Interpretation of Extensive and Intensive Margin Effects in Generalized Tobit Models, Kevin E. Staub,33 p.
1011	Impact of Specialization on Health Outcomes – Evidence from U.S. Cancer Data, Johannes Schoder, Frank R. Lichtenberg,19 p.
1010	Renewable energy policy in the presence of innovation: does government pre-commitment matter?, Reinhard Madlener, Ilja Neustadt, September 2010, 42 p.
1009	Do Religious Beliefs Explain Preferences for Income Redistribution? Experimental Evidence, Ilja Neustadt, September 2010, 36 p.
1008	Lobbying and the Power of Multinational Firms, Andreas Polk, Armin Schmutzler, Adrian Muller, August 2010, 32 p.
1007	When Are Preferences Consistent? The Effects of Task Familiarity and Contextual Cues on Revealed and Stated Preferences, Felix Schläpfer, Baruch Fischhoff, August 2010, 20 p.
1006	Golden Balls: A Prisoner's Dilemma Experiment, Donja Darai, Silvia Grätz, July 2010, 47 p.
1005	Probability Weighting as Evolutionary Second-best, Florian Herold, Nick Netzer, July 2010, 32 p.
1004	Trade Openness, Gains from Variety and Government Spending, Sandra Hanslin, April 2010, 46 p.
1003	Is the Welfare State Sustainable? Experimental Evidence on Citizens' Preferences for Income Redistribution, Ilja Neustadt, Peter Zweifel, March 2010, 32 p.
1002	Preferences for Health Insurance in Germany and the Netherlands – A Tale of Two Countries, Peter Zweifel, Karolin Leukert, Stephanie Berner, March 2010, 22 p.
1001	Convex Treatment Response and Treatment Selection, Stefan Boes, January 2010, 28 p.
0920	Bounds on Counterfactual Distributions Under Semi-Monotonicity Constraints, Stefan Boes, December 2009, 38 p.
0919	Rotten Kids with Bad Intentions, Nick Netzer, Armin Schmutzler, December 2009, 38 p.
0918	Partial Identification of Discrete Counterfactual Distributions with Sequential Update of Information, Stefan Boes, December 2009, 37 p.
0917	How much do journal titles tell us about the academic interest and relevance of economic reserach? An empirical analysis. Felix Schläpfer, December 2009, 14 p.
0916	Fine Tuning of Health Insurance Regulation: Unhealthy Consequences for an Individual Insurer, Johannes Schoder, Michèle Sennhauser, Peter Zweifel, August 2009, 18 p.
0915	Capping Risk Adjustment?, Patrick Eugster, Michèle Sennhauser, Peter Zweifel, September 2009, 27 p.
0914	A Pharmaceutical Innovation: Is it Worth the Money? Whose Money?, Michèle Sennhauser, Peter Zweifel, September 2009, 22 p.
0913	Copula-based bivariate binary response models, Rainer Winkelmann, August 2009, 26 p.
0912	Simulating WTP Values from Random-Coefficient Models, Maurus Rischatsch, July 2009, 6 p.

0911 Physician dispensing and the choice between generic and brand-name drugs – Do margins affect choice?, Maurus Rischatsch, Maria Trottmann, July 2009, 15 p.
0910 GPs' preferences: What price fee-for-service?, Peter Zweifel, Maurus Rischatsch, Angelika Brändle, July 2009, 21 p.
0909 Social Mobility and Preferences for Income Redistribution: Evidence from a Discrete Choice Experiment, Ilja Neustadt, Peter Zweifel, July 2009, 31 p.
0908 Robust estimation of zero-inflated count models, Kevin E. Staub, Rainer Winkelmann June 2009, 22 p.
0907 Competitive Screening in Insurance Markets with Endogenous Wealth Heterogeneity, Nick Netzer, Florian Scheuer, April 2009, 28 p.
0906 New Flight Regimes and Exposure to Aircraft Noise: Identifying Housing Price Effects Using a Ratio-of-Ratios Approach, Stefan Boes, Stephan Nüesch, April 2009, 40 p.
0905 Patents versus Subsidies – A Laboratory Experiment, Donja Darai, Jens Großer, Nadja Trhal, March 2009, 59 p.
0904 Simple tests for exogeneity of a binary explanatory variable in count data regression models, Kevin E. Staub, February 2009, 30 p.
0903 Spurious correlation in estimation of the health production function: A note, Sule Akkoyunlu, Frank R. Lichtenberg, Boriss Siliverstovs, Peter Zweifel, February 2009, 13 p.
0902 Making Sense of Non-Binding Retail-Price Recommendations, Stefan Bühler, Dennis L. Gärtner, February 2009, 30 p.
0901 Flat-of-the-Curve Medicine – A New Perspective on the Production of Health, Johnnes Schoder, Peter Zweifel, January 2009, 35 p.
0816 Relative status and satisfaction, Stefan Boes, Kevin E. Staub, Rainer Winkelmann, December 2008, 11 p.
0815 Delay and Deservingness after Winning the Lottery, Andrew J. Oswald, Rainer Winkelmann, December 2008, 29 p.
0814 Competitive Markets without Commitment, Nick Netzer, Florian Scheuer, November 2008, 65 p.
0813 Scope of Electricity Efficiency Improvement in Switzerland until 2035, Boris Krey, October 2008, 25 p.
0812 Efficient Electricity Portfolios for the United States and Switzerland: An Investor View, Boris Krey, Peter Zweifel, October 2008, 26 p.
0811 A welfare analysis of "junk" information and spam filters; Josef Falkinger, October 2008, 33 p.
0810 Why does the amount of income redistribution differ between United States and Europe? The Janus face of Switzerland; Sule Akkoyunlu, Ilja Neustadt, Peter Zweifel, September 2008, 32 p.
0809 Promoting Renewable Electricity Generation in Imperfect Markets: Price vs. Quantity Policies; Reinhard Madlener, Weiyu Gao, Ilja Neustadt, Peter Zweifel, July 2008, 34p.
0808 Is there a U-shaped Relation between Competition and Investment? Dario Sacco, July 2008, 26p.
0807 Competition and Innovation: An Experimental Investigation, May 2008, 20 p.
0806 All-Pay Auctions with Negative Prize Externalities: Theory and Experimental Evidence, May 2008, 31 p.
0805 Between Agora and Shopping Mall, Josef Falkinger, May 2008, 31 p.